Mon. Not. R. Astron. Soc. **321**, 471–474 (2001)

The dark matter problem in disc galaxies

James Binney,[1★] Ortwin Gerhard[2] and Joseph Silk[1]
[1]*Physics Department, Oxford University, Keble Road, Oxford, OX1 3NP*
[2]*Astronomisches Institut, Universität Basel, Venusstrasse 7, CH-4102 Binningen, Switzerland*

Accepted 2000 September 12. Received 2000 July 3; in original form 2000 March 13

ABSTRACT

In the generic CDM cosmogony, dark-matter haloes emerge too lumpy and centrally concentrated to host observed galactic discs. Moreover, discs are predicted to be smaller than those observed. We argue that the resolution of these problems may lie with a combination of the effects of protogalactic discs, which would have had a mass comparable to that of the inner dark halo and be plausibly non-axisymmetric, and of massive galactic winds, which at early times may have carried off as many baryons as a galaxy now contains. A host of observational phenomena, from quasar absorption lines and intracluster gas through the G-dwarf problem, point to the existence of such winds. Dynamical interactions will homogenize and smooth the inner halo, and the observed disc will be the relic of a massive outflow. The inner halo expanded after absorbing energy and angular momentum from the ejected material. Observed discs formed at the very end of the galaxy formation process, after the halo had been reduced to a minor contributor to the central mass budget and strong radial streaming of the gas had died down.

Key words: galaxies: formation – cosmology: theory.

1 INTRODUCTION

High resolution simulations of galaxy formation, incorporating realistic cold dark matter (CDM) initial conditions of dark halo formation, generally confirm the existence of a universal density (NFW) profile in the outer regions of galaxies (Navarro, Frenk & White 1997). Moreover, some groups are now reporting significant central dark matter density cusps that are as steep as $\propto r^{-\beta}$ with $\beta \approx 1.5$. The best observed low surface brightness galaxies, which are dark matter dominated, have rotation curves that are inconsistent with β values as steep as 1.5 (van den Bosch & Swaters 2000).

Two further problems encountered with the cold dark matter hypothesis are (i) that the scale-lengths of discs are predicted to be too small by a factor ~ 5 (Steinmetz & Navarro 1999) and (ii) an order of magnitude more satellites are predicted than are observed (Moore et al. 1999). Both of these problems are closely related to the persistence of substructure in high-resolution *N*-body simulations of hierarchical models of dark halo formation.

There are two possible avenues for resolution of these problems. One approach is to tinker with the particle physics. One may abandon the idea that CDM is weakly interacting. There are CDM particle candidates for which annihilation rates are of the order of the weak rate but for which scattering cross-sections are of the order of the strong interaction (Carlson, Machacek & Hall 1992; Machacek & Hall 1992). Such dissipative CDM may erase both the CDM cusps and clumpiness (Spergel & Steinhardt 2000), but at the price of introducing an unacceptably spherical inner core in massive clusters (Miralda-Escude 2000). One may suppress the small-scale power on subgalactic scales, either by invoking broken scale-invariance (Kamionkowski & Liddle 1999) or warm dark matter (Sommer-Larsen & Dolgov 1999), in the hope that the structure of massive dark haloes will be modified.

Here we adopt the less radical approach of exploring astrophysical alternatives. We accept the fundamental correctness of the CDM picture, and ask (i) could excess dark matter be ejected from the optical galaxy and (ii) why do baryons in galaxies currently have more specific angular momentum than predicted by the simple CDM picture? We argue that these questions are connected, and that both may be resolved if galaxies have first absorbed and then ejected a mass of baryons which is comparable to their current baryonic masses. An earlier paper argued that baryonic winds can imprint cores within dwarf galaxy dark haloes (Navarro, Eke & Frenk 1996). Here we propose that energy and angular momentum surrendered by the ejected baryons have profoundly modified the dark halo within the current optical massive galaxy. In this picture most protogalactic material remained gaseous until the period of mass ejection was substantially complete – this conjecture is tenable because we have no reliable knowledge of either the rate at which, or the efficiency with which, stars form in a protogalactic environment.

In Section 2 we argue for massive galactic outflows. In Section 3 we ask how the dark halo was modified as a result of processing the material prior to ejection. Section 4 is concerned with the

★ E-mail: binney@thphys.ox.ac.uk

implications for the star-formation rate within a gaseous bar. Section 5 sums up our arguments.

2 INFLOW AND OUTFLOW

A primary problem with the conventional picture of galaxy formation is that in all simulations, the baryons lose much of their angular momentum as they fall into dark-matter haloes (Katz & Gunn 1991; Navarro & Benz 1991; Weil, Eke & Efstathiou 1998) rather than conserving it as semi-analytic models of galaxy formation typically assume (Kauffmann, White & Guiderdoni 1993; Granato et al. 2000). Consequently, whereas in semi-analytic galaxy-formation models, the baryons are marginally short of angular momentum to account for the observed disc sizes, in reality they will fall short by an order of magnitude.

Current estimates of the acquisition of angular momentum by perturbations in the expanding universe seem robust, as is the prediction that collapsing baryons will surrender much of their angular momentum. Hence, we should take seriously the expectation that protogalaxies early on will contain a considerable mass of low angular momentum baryons. The mass in question is comparable to the masses of observed discs, but little of this material can end up in the final disc, because it is short of the required angular momentum by a factor of several. So where *does* this material end up? Some of it will have been converted into the galaxy's bulge and central black hole. However, the bulge and central black hole of the Milky Way between them contain less mass than the disc by a factor ~ 5. In galaxies of later Hubble type, such as M33, the factor can be substantially greater.

Star formation is always associated with conspicuous outflows, which are thought to be generically associated with accretion discs. Hence, it is likely that the majority of a protogalaxy's low angular momentum baryons are ejected in a wind that is powered by star formation, magnetic torques and black hole accretion. Observations of starburst galaxies such as M82 lend direct support to this conjecture (Axon & Taylor 1978; Dahlem, Weaver & Heckman 1998) as does the work of Pettini et al. (2000) on the gravitationally lensed galaxy MS 1512-cB58 at $z = 2.7268$. We conclude that the observed discs of galaxies plausibly formed from the higher angular momentum tail of the conventional distribution. In terms of a spherically-symmetric infall model, we imagine that the baryons, which started out close to the centre of the protogalaxy, were mostly ejected. Galactic discs were formed from the baryons which were initially confined to the periphery of the volume inside the halo virial radius, from which the final galaxy's dark matter was drawn. Because dissipative material is much better able to find its way down to minima of the gravitational potential than dark matter, a significant fraction of the baryons currently in galaxies could come from this region. On account of its large initial galactocentric radii, this material will have started out with *higher* specific angular momentum than the disc into which it was destined to settle.

X-ray observations of early-type galaxies and clusters of galaxies provide strong support for the conjecture that forming galaxies blow powerful winds. In clusters of galaxies, the metal-enriched ejecta are directly observed because they have been trapped by the cluster potential (Arnaud et al. 1992; Loewenstein & Mushotzky 1996). The wide spread in the X-ray luminosities of the hot atmospheres of giant elliptical galaxies has been used to argue persuasively that early winds can escape the potentials of many galaxies, but not those of the most massive systems, with the consequence that the ejected gas sometimes falls back into the visible galaxy and gives rise to a 'cooling flow' (D'Ercole et al. 1989). Early heating of intergalactic gas by galactic winds has been invoked to explain the steepness of the observed X-ray luminosity–temperature relation (Kaiser 1991).

Independent arguments point to massive outflows early in galactic evolution. The narrow dispersion in the colour-magnitude diagrams of cluster ellipticals, both now and at redshifts $z \sim 1$ (Jorgensen et al. 1999), implies that the galaxies' colours are not heavily contaminated by metal-poor stars. Early outflows would prevent such contamination (Kauffmann & Charlot 1998; Ferreras & Silk 2000). Moreover, bulges and the nuclei of elliptical galaxies are enhanced in α-elements (C, O, Mg) relative to Fe (Kuntschner 2000). This observation seems to require suppression of star formation from material that has been enriched in Fe by Type Ia supernovae. It is often assumed that this suppression is achieved by converting all the protogalactic gas to stars before many Type Ia supernovae have exploded, but it could also be achieved by a supernova-driven wind which removed the remaining gas before the Type Ia explosions. Models of the chemical evolution of discs (Prantzos & Silk 1998) similarly yield an acceptably small number of metal-poor stars in the old disc if a supernova-driven wind carries metal-enriched gas out of the galaxy. Finally, the detection of old halo white dwarfs with a frequency and mass range similar to that inferred for massive compact halo objects from the Large Magellanic Cloud (LMC) microlensing experiments (Ibata et al. 1999) will, if spectroscopically confirmed (Ibata et al. 2000), require a substantial protogalactic outflow phase to eliminate from the protogalaxy heavy elements which would otherwise pollute stars that formed later and are observed to have low metallicities.

Galactic outflows will have delivered heavy elements to the intergalactic medium (Lehnert, Heckman & Weaver 1999). This process not only accounts for the observed metallicities of intracluster gas (Renzini 1997), but may also be responsible for the metallicities of the low-density gas which is primarily detected through Lyα absorption in quasar spectra. It is possible that at low z, significant enrichment of the intergalactic medium (IGM) and intracluster medium (ICM) might come from dwarf galaxies, although the low metallicities of the dwarfs argue against this unless the luminosity function is exceptionally steep. At the redshifts $z \gtrsim 2$ at which the enriched IGM material is observed, most of the stars in the nearby dwarf galaxies will not have formed, so the more luminous galaxies would have necessarily had to dominate unless a new population of early-forming dwarfs is invoked. However semi-analytic theory predicts that at $z \gtrsim 2$, most star formation is confined to locations at which luminous galaxies now reside (Baugh et al. 1998; Benson et al. 2000). These locations are far removed from the low-density gas, which is observed to contain heavy elements. Galactic winds could be responsible for transporting the heavy elements from the location of the bulk of star formation, to where they are observed. Moreover, extended metal-enriched absorption systems might arise from expanding shells which form in galactic winds in the same way that shells form around planetary nebulae.

Thus, many lines of argument suggest that outflows from both spheroids and discs were common, and therefore that significantly more baryons were involved in the formation of a given galaxy than it now contains.

3 MODIFICATION OF THE HALO

As we have seen, the infalling baryons will have lost much of their angular momentum. The lost angular momentum is taken up by the halo. In principle, acquisition of this angular momentum modifies the halo at all radii, but the modifications are small where dark matter dominates over baryons, and are profound only interior to the radius at which $M_{\rm disc} \sim M_{\rm halo}$. Observationally, we know that the baryons are dominant inside the solar radius, so we expect the halo profile to be substantially modified there, precisely as the CDM model seemingly requires (Navarro & Steinmetz 2000).

There are three obvious mechanisms by which gas can lose angular momentum to the halo. Early on, the halo is expected to be triaxial and its principal axes will rotate slowly if at all. Gas flowing in such a potential rapidly loses angular momentum, even if its mass is small compared to the mass of the local halo (Katz & Gunn 1991). If gas ever accumulates to the degree that it contributes a non-negligible fraction of the mass interior to some radius r, two other mechanisms for angular-momentum loss become effective: massive blobs of gas will lose angular momentum through dynamical friction (Stark et al. 1991; Navarro & Steinmetz 1997), and a tumbling gaseous bar will lose angular momentum through resonant coupling (Hernquist & Weinberg 1995). These last two processes operate even if the halo becomes axisymmetric, as it may do where gas contributes significantly to the overall mass budget.

During the earliest stages of galaxy formation, gas will be far from centrifugal equilibrium and will flow rapidly inwards. If massive clumps form during this period, they will quickly fall further into the centre, losing both energy and angular momentum. If massive clumps cannot form rapidly, the gas loses energy faster than angular momentum, with the consequence that gas which started out at a given galactocentric radius will eventually settle to a (possibly elliptical) ring. If this ring is not substantially self-gravitating, it will evolve little if at all. Present-day low surface brightness galaxies would seem to be made up of such inert rings of gas.

If the ring is significantly self-gravitating, it will continue to lose angular momentum to the local halo by a combination of dynamical friction and bar-driven resonant coupling. The dynamics of a tumbling gaseous bar embedded in a dark halo of comparable mass has yet to be carefully studied, but in phase space, orbits at energies around the bar's corotation energy will be highly chaotic, and the strong orbital shear, which is characteristic of chaos, will tend to erase substructure within the halo near the corotation energy. Both analytic calculations and simulations show that, in the case of a stellar bar, resonant coupling is a rapid process: the time-scale of angular-momentum loss exceeds the bar's dynamical time by a factor of only a few (Weinberg & Tremaine 1984; Debattista & Sellwood 1998). Consequently, a bar embedded in a dynamically significant halo will shrink. This shrinkage will rapidly enhance the mass fraction of baryons. Moreover, concentration of the baryons will be accompanied by expansion of the local halo as it takes up energy and angular momentum shed by the bar.

The coupling between baryons and dark matter is a fairly local process, essentially confined to a factor of 2 either side of the baryons' corotation radius. The processes we have described for one corotation radius presumably occurred in sequence at a series of radii which increased from very small values out to scales characteristic of present-day disc galaxies. If the arguments of the preceeding section are correct, the dark matter at any given radius r will interact locally with many different parcels of baryons during the formation process, as these parcels move through radius r on their way to the galactic centre and probable ejection from the galaxy.

These considerations suggest that, if the baryons ever become dynamically significant, they will go on losing angular momentum to the halo until they are dominant, and that dominance is achieved by a combination of the baryons moving in and the dark matter moving out, accompanied by the expulsion of a significant fraction of the baryonic mass in early winds. An extended phase of baryonic infall may then follow; chemical evolution models of the Milky Way disc require about half of the disc to have formed via late infall (Prantzos & Silk 1998). Late infall may double the mass of the disc, with the consequence that the final disc is close to maximal, and the role of dark matter within the solar circle is negligible.

4 BARS AND STAR FORMATION

Since the stars of the current disc are now on nearly circular orbits, they cannot have formed until after any tumbling gaseous bar had dissolved. Is it reasonable to have a bar without significant star formation? The dwarf galaxy NGC 2915 (Bureau et al. 1999) is an example of a dark-matter dominated galaxy with a very extended HI disc revealing a central bar and spiral structure extending well beyond the optical component. Evidently the Toomre Q of this system satisfies $Q_{\rm global} > Q > Q_{\rm local}$, where $Q_{\rm global}$ and $Q_{\rm local}$ are the critical values of the disc instability parameter for global non-axisymmetric and local axisymmetric instabilities, respectively. One can readily imagine that as the disc forms, the gas surface density increases and the gas velocity dispersion drops, so that Q decreases, and the local Q criterion is subsequently satisfied. In the solar neighbourhood the disc satisfies

$$Q_{\rm local} \approx \left(\frac{\sigma_{\rm g}}{10\,{\rm km\,s^{-1}}}\right)\left(\frac{15\,{\rm M_\odot\,pc^{-2}}}{\mu_{\rm g}}\right)$$

and is marginally unstable. The gas disc of the Milky Way presently contains about $6\times10^9\,{\rm M_\odot}$. In the transient bar phase, the effective Q is increased by the ratio of bar streaming velocity to gas velocity dispersion $\sigma_{\rm g}$, which amounts to a factor of $\sim$10. Hence a gas mass of up to $\sim 10^{11}\,{\rm M_\odot}$ can be stabilized against star formation during the transient bar phase.

It is clear that high resolution numerical simulations are required to model the coupling between the non-axisymmetric protodisc and the dark halo. These simulations need to include the effects of baryonic dissipation and star formation. There may be possible stellar relics of an early massive bar, which would be recognizable as a disc component of old stars with significant orbital eccentricity.

5 CONCLUSIONS

Two serious problems currently plague the CDM theory of galaxy formation: an excess of dark matter within the optical bodies of galaxies, and discs that are too small. The second problem reflects the low angular momentum of infalling matter, and is made worse when one accepts that infalling baryons will surrender much of their angular momentum to the dark halo. In consequence, galaxies start with more low angular momentum baryons than they currently hold in their bulges and central black holes. We have argued that the surplus material was early on ejected as a massive wind. Many direct and indirect observational arguments point to the existence of such winds.

Although the angular momentum of the first baryons to fall in was inadequate for the formation of the disc, it was not entirely negligible, and caused the inner halo to expand when the latter absorbed it. Similarly, the angular momentum of the baryons, which are now in the disc, was originally larger than it now is, and the surplus angular momentum further expanded the inner halo. In short, through relieving perhaps twice the baryonic mass of the current galaxy of angular momentum and energy, the dark halo has become substantially less centrally concentrated than it was originally, and it now contributes only a small fraction of the mass within the visible galaxy. During this refashioning of its inner parts, substructure is likely to have been erased, leaving the final inner halo smoother both locally and globally.

This picture requires the baryonic mass to remain gaseous until the dark halo has been reduced to a minor contributor to the central mass, and a disc has formed in which most material is on circular orbits. This conjecture is plausible for two reasons: (i) the dark halo will be unresponsive to the collective modes of a gaseous disc, so the disc will not have growing modes until it dominates the gravitational potential in which it sits, and (ii) the enhanced orbital shear that is characteristic of closed orbits in a barred potential cannot be conducive to star formation except near the bar's centre. In any case we have little understanding of what controls the rate of star formation in a protogalaxy, and we know from the fragility of discs (Toth & Ostriker 1992) that discs formed at the end of the formation process, after merging, had all but ceased and the largest substructure had been erased from bulge and inner halo.

Existing numerical simulations of the interactions of baryons and dark matter during galaxy formation (e.g. Navarro & Steinmetz 2000; Benson et al. 2000) lack both the mass resolution and some of the physics which is required to realise the essential ideas employed here. For example, in the simulations of Benson et al. the gravitational softening length is $10\,h^{-1}$ kpc, and the basic baryonic resolution element has mass $\sim 4 \times 10^{10}\,\mathrm{M}_{\odot}$ and spurious discreteness effects will be present on mass scales several times larger. Such simulations neglect magnetic fields (which are believed to drive winds off accretion discs) and energy input by both supernovae and the central massive object.

In summary, a considerable mass of low angular momentum baryons must have been ejected. This prediction is a priori plausible, given observations of winds from starburst galaxies and outflows from Lyman-break galaxies, and given the prevalence of outflows in star-formation regions. The heavy element abundances of hot gas in clusters of galaxies and in cool, low-density gas, observed at redshifts $z \sim 2$ through quasar absorption lines, are likely to arise through the mixing of metal-rich ejecta with primordial gas. The low angular momentum material having been ejected, the current discs formed from the higher angular momentum baryons which fell in later. Since the ejection stage commences only once $M_{\mathrm{baryon}}(r) \sim M_{\mathrm{dm}}(r)$, when self-gravity can drive gas flows and the ensuing winds, thereby causing the dark-matter distribution to expand and the baryons to contract further, the visible galaxy is inevitably baryon dominated and yet has a circular speed which approximately matches that of the embedding halo. Thus the so-called 'disc–halo conspiracy' (Bahcall & Casertano 1995) is not really a coincidence but a consequence of dynamical evolution.

REFERENCES

Arnaud M., Rothenflug R., Boulade O., Vigroux L., Vangioni-Flan E., 1992, A&A, 254, 49
Axon D. J., Taylor K., 1978, Nat, 274, 37
Bahcall J., Casertano S., 1995, ApJ, 293, L7
Baugh C. M., Cole S., Frenk C. S., Lacey C. G., 1998, ApJ, 498, 504
Benson A. J., Pearce F. R., Frenk C. S., Baugh C. M., Jenkins A., 2000, MNRAS, in press, astro-ph 9912220
van den Bosch F., Swaters R., 2000, AJ, submitted, astro-ph 0006048
Bureau M., Freeman K. C., Pfitzner D. W., Meurer G. R., 1999, AJ, 118, 2158
Hernquist L., Weinberg M., 1995, ApJ, 400, 80
Carlson E., Machacek M., Hall L., 1992, ApJ, 398, 43
Dahlem M., Weaver K. A., Heckman T. M., 1998, ApJS, 118, 401
Debattista V. P., Sellwood J. A., 1998, ApJ, 493, L5
D'Ercole A., Renzini A., Ciotti L., Pellegrini S., 1989, ApJ, 341, L9
Fereras I., Silk J., 2000, ApJ, in press, preprint astro-ph/9910385
Granato G. L., Lacey C. G., Silva L., Bressan A., Baugh C. M., Cole S., Frenk C. S., 2000, ApJ, 542, 710
Ibata R. A., Richer H., Gilliland R., Scott D., 1999, ApJ, 524, L1
Ibata R. A., Irwin M., Bienayme O., Scholz R., Guibert J., 2000, ApJ, 532, L41
Jorgensen I., Franx M., Hjorth J., van Dokkum P. G., 1999, MNRAS, 308, 833
Kaiser N., 1991, ApJ, 383, 104
Kamionkowski M., Liddle A., 1999, Phys. Rev. Lett., 84, 4525
Katz N., Gunn J. E., 1991, ApJ, 377, 365
Kauffmann G., Charlot S., 1998, MNRAS, 294, 705
Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201
Kuntschner H., 2000, MNRAS, 315, 184
Lehnert M. D., Heckman T. M., Weaver K. A., 1999, ApJ, 523, 575
Loewenstein M., Mushotzky R. F., 1996, ApJ, 466, 695
Machacek M., Hall L., 1992, ApJ, 431, 41
Miralda-Escude J., 2000, preprint astro-ph/0002050
Moore B., Ghigna S., Governato F., Luke G., Quinn T., Stadel J., Tozzi P., 1999, ApJ, 524, 19
Navarro J. F., Benz W., 1991, ApJ, 380, 320
Navarro J. F., Eke V. R., Frenk C., 1996, MNRAS, 283, L72
Navarro J. F., Frenk C., White S. D. M., 1997, ApJ, 490, 493
Navarro J. F., Steinmetz M., 1997, ApJ, 478, 13
Navarro J. F., Steinmetz M., 2000, ApJ, 528, 607
Pettini M., Steidel C. C., Adelberger K. L., Dickinson M., Giavalisco M., 2000, ApJ, 528, 96
Prantzos N., Silk J., 1998, ApJ, 507, 229
Renzini A., 1997, ApJ, 488, 35
Sommer-Larsen J., Dolgov A., 1999, preprint astro-ph/9912166
Spergel D. N., Steinhardt P., 2000, Phys. Rev. Lett., 84, 3760
Stark A. A., Gerhard O.E, Binney J. J., Bally J., 1991, MNRAS, 248, 14P
Steinmetz M., Navarro J. F., 1999, ApJ, 513, 555
Toth G., Ostriker J. P., 1992, ApJ, 389, 5
Weil M., Eke V. R., Efstathiou G., 1998, MNRAS, 300, 773
Weinberg M., Tremaine S., 1984, MNRAS, 209, 729

This paper has been typeset from a TEX/LATEX file prepared by the author.